AUTHENTICATED
U.S. GOVERNMENT
INFORMATION
GPO

that is made up of one or more noncontiguous segments of a larger sequence or segments from different sequences shall be presented as a separate sequence.

[63 FR 29635, June 1, 1998]

§ 1.823 Requirements for nucleotide and/or amino acid sequences as part of the application papers.

(a) The ''Sequence Listing'' required by § 1.821(c), setting forth the nucleotide and/or amino acid sequences and associated information in accordance with paragraph (b) of this section, must begin on a new page and must be titled ''Sequence Listing''. The ''Sequence Listing'' preferably should be numbered independently of the numbering of the remainder of the application. Each page of the ''Sequence Listing'' should contain no more than 66 lines and each line should contain no more than 72 characters. A fixed-width font should be used exclusively throughout the ''Sequence Listing.''

(b) The ''Sequence Listing'' shall, except as otherwise indicated, include the actual nucleotide and/or amino acid sequence, the numeric identifiers and their accompanying information as shown in the following table. The numeric identifier shall be used only in the ''Sequence Listing.'' The order and presentation of the items of information in the ''Sequence Listing'' shall conform to the arrangement given below. Each item of information shall begin on a new line and shall begin with the numeric identifier enclosed in angle brackets as shown. The submission of those items of information designated with an ''M'' is mandatory. The submission of those items of information designated with an ''O'' is optional. Numeric identifiers <110> through <170> shall only be set forth at the beginning of the ''Sequence Listing.'' The following table illustrates the numeric identifiers.

Numeric identifier	Definition	Comments and format	Mandatory (M) or optional (O).
<110>	Applicant	Preferably max. of 10 names; one name per line; preferable format: Surname, Other Names and/or Initials.	M.
<120>	Title of Invention		M.
<130>	File Reference	Personal file reference	M when filed prior to assignment of appl. number.
<140>	Current Application Number.	Specify as: US 07/999,999 or PCT/US96/99999.	M, if available.
<141>	Current Filing Date	Specify as: yyyy–mm–dd	M, if available.
<150>	Prior Application Number.	Specify as: US 07/999,999 or PCT/US96/99999.	M, if applicable include priority documents under 35 USC 119 and 120.
<151>	Prior Application Filing Date.	Specify as: yyyy–mm–dd	M, if applicable.
<160>	Number of SEQ ID NOs.	Count includes total number of SEQ ID NOs.	M.
<170>	Software	Name of software used to create the Sequence Listing.	O.
<210>	SEQ ID NO:#:	Response shall be an integer representing the SEQ ID NO shown.	M.
<211>	Length	Respond with an integer expressing the number of bases or amino acid residues.	M.
<212>	Type	Whether presented sequence molecule is DNA, RNA, or PRT (protein). If a nucleotide sequence contains both DNA and RNA fragments, the type shall be ''DNA.'' In addition, the combined DNA/RNA molecule shall be further described in the <220> to <223> feature section.	M.
<213>	Organism	Scientific name, i.e. Genus/ species, Unknown or Artificial Sequence. In addition, the ''Unknown'' or ''Artificial Sequence'' organisms shall be further described in the <220> to <223> feature section.	M

Numeric identifier	Definition	Comments and format	Mandatory (M) or optional (O).
<220>	Feature	Leave blank after <220>. <221–223> provide for a description of points of biological significance in the sequence..	M, under the following conditions: if ''n,'' ''Xaa,'' or a modified or unusual L-amino acid or modified base was used in a sequence; if ORGANISM is ''Artificial Sequence'' or ''Unknown''; if molecule is combined DNA/RNA''
<221>	Name/Key	Provide appropriate identifier for feature, preferably from WIPO Standard ST.25 (1998), Appendix 2, Tables 5 and 6.	M, under the following conditions: if ''n,'' ''Xaa,'' or a modified or unusual L-amino acid or modified base was used in a sequence.
<222>	Location	Specify location within sequence; where appropriate state number of first and last bases/amino acids in feature.	M, under the following conditions: if ''n,'' ''Xaa,'' or a modified or unusual L-amino acid or modified base was used in a sequence.
<223>	Other Information	Other relevant information; four lines maximum.	M, under the following conditions: if ''n,'' ''Xaa,'' or a modified or unusual L-amino acid or modified base was used in a sequence; if ORGANISM is ''Artificial Sequence'' or ''Unknown''; if molecule is combined DNA/RNA.
<300>	Publication Information.	Leave blank after <300>	O.
<301>	Authors	Preferably max of ten named authors of publication; specify one name per line; preferable format: Surname, Other Names and/or Initials.	O.
<302>	Title		O.
<303>	Journal		O.
<304>	Volume		O.
<305>	Issue		O.
<306>	Pages		O.
<307>	Date	Journal date on which data published; specify as yyyy–mm–dd, MMM–yyyy or Season-yyyy.	O.
<308>	Database Accession Number.	Accession number assigned by database including database name.	O.
<309>	Database Entry Date.	Date of entry in database; specify as yyyy–mm–dd or MMM–yyyy.	O.
<310>	Patent Document Number.	Document number; for patent-type citations only. Specify as, for example, US 07/999,999.	O.
<311>	Patent Filing Date	Document filing date, for patent-type citations only; specify as yyyy–mm–dd.	O.
<312>	Publication Date	Document publication date, for patent-type citations only; specify as yyyy–mm–dd.	O.
<313>	Relevant Residues	FROM (position) TO (position)	O.
<400>	Sequence	SEQ ID NO should follow the numeric identifier and should appear on the line preceding the actual sequence.	M.

[63 FR 29636, June 1, 1998]

§ 1.824 Form and format for nucleotide and/or amino acid sequence submissions in computer readable form.

(a) The computer readable form required by §1.821(e) shall meet the following specifications:

(1) The computer readable form shall contain a single ''Sequence Listing'' as either a diskette, series of diskettes, or other permissible media outlined in paragraph (c) of this section.

(2) The ''Sequence Listing'' in paragraph (a) (l) of this section shall be submitted in American Standard Code for Information Interchange (ASCII) text. No other formats shall be allowed.

(3) The computer readable form may be created by any means, such as word processors, nucleotide/amino acid sequence editors or other custom computer programs; however, it shall conform to all specifications detailed in this section.

(4) File compression is acceptable when using diskette media, so long as the compressed file is in a self-extracting format that will decompress on one of the systems described in paragraph (b) of this section.